



University of Maryland School of Public Health

EPIB660 – Analysis of National Health Survey Data

Semester: Summer 2018
Classroom and Time: N/A – online course
Course Website: See ELMS/Canvas
Instructor: Dr. Taylor Lewis
Instructor Website: <http://mason.gmu.edu/~tlewis18/>
Office Hours: By appointment
Phone: 703-517-3563
Email: tlewis18@gmu.edu

Course Description:

The purpose of this course is to provide background on how features such as stratification, clustering, and unequal sample selection probabilities can invalidate the assumptions underlying traditional statistical techniques, those implicitly assuming a simple random sampling with replacement design. This is not a course on sample design. Rather, it is a course on the secondary analysis of survey data sets derived from a sample design involving one of the aforementioned features. The formulaic modifications these features introduce are demonstrated for a wide range of statistical analyses using the SURVEY family of SAS/STAT® procedures (Version 9.4 or later). Examples in class and assignments give students exposure to analyzing publically available data produced from nationally representative health surveys such as the National Health Interview Survey (NHIS), the National Immunization Survey (NIS), and the National Health and Nutritional Examination Survey (NHANES).

Course Pre-Requisites:

EPIB650 Biostatistics I. EPIB697 Public Health Data Management is highly recommended. Or, permission from instructor.

Course Learning Objectives:

Upon completing this course, the student will be able to:

1. Understand the terminology and rationale of sampling techniques commonly used in applied survey research and why simple random sampling is the exception rather than the rule.
2. Learn how to consult documentation of publically available health survey data sets to identify whether features such as stratification, clustering, and unequal weights are present.
3. Confidently account for these features in health survey data by making use of the SURVEY family of SAS procedures for a wide range of estimation and analysis tasks, including a variety of regression models.
4. Understand how to properly analyze subsets of a survey data set for the analysis of population subgroups.

Program Competencies Addressed in this Course:

The following competencies are addressed in this course:

1. Describe the role biostatistics serves in the discipline of public health.
2. Select appropriate inferential statistical methods to answer research questions relevant to public health research.
3. Conduct descriptive and inferential statistical analyses that are appropriate to different basic study designs used in public health research.

4. Identify statistical approaches to address threats to validity in epidemiologic studies.
5. Interpret results of statistical analyses found in public health studies.
6. Critically review and summarize statistical analyses presented in public health literature.

Required Text:

Lewis, T. (2016). *Complex Survey Data Analysis with SAS®*. Boca Raton, FL: CRC Press. ISBN 9781498776776.

Required Software:

This course will make exclusive use of SAS software. If you do not already have access via your school department or employer, perhaps the best option is to look into acquiring a free license of SAS University Edition: http://www.sas.com/en_us/software/university-edition.html. There is also a free, Web-based version called SAS OnDemand for Academics: https://www.sas.com/en_us/software/on-demand-for-academics.html.

Course Website:

Audio lectures, lecture notes, data sets, homework assignments, and important course-related announcements will be posted to the course Canvas page from within ELMS (Enterprise Learning Management System). Please check it on a regular basis. You may wish to print the lecture notes prior to viewing each audio recording so that you can annotate as necessary.

You can access the course website by following these directions:

- Open a Web browser and go to <https://myelms.umd.edu/login>
- Enter your Directory ID and Password
- Click “Courses” on the ELMS home tab
- Click “EPIB660 Analysis of Public Health Survey Data”

Course Requirements and Expectations:

Lectures: Material to be covered on assignments and the final exam will be presented over a sequence of ten modules. For best performance, students should complete the required readings prior to watching the recorded lecture associated with a module. If a student misses a module, he/she may be missing exposure to key concepts. The PowerPoint slides used in the recorded lectures are available for download from the course website.

Homework Assignments: There will be five homework assignments, each designed to give students practice applying concepts presented in two modules at a time (see detailed course outline below) on real-world public survey data. Late homework will NOT be accepted without a reasonable and advance notice. If desired, you may submit scanned handwritten solutions to homework assignments—in part or in whole—but no matter the format, please ensure your homework assignment is uploaded as a single Word or PDF file with your answers clearly labeled.

Final Exam: There will be a take-home final exam made available online via the course website at 6:00 a.m. on Tuesday 8/14/18. The final exam must be submitted by 11:59 p.m. on Sunday 8/19/18. It will be open book and open notes. As a general rule, a make-up exam or advance exam will NOT be given. Exceptions to this rule will be evaluated on a case-by-case basis.

University Course Related Policies:

All University of Maryland-approved course policies are provided at the following website:

<http://www.ugst.umd.edu/courserelatedpolicies.html>

Policy descriptions, resources, and links to official policy documents are provided for:

- Academic Integrity: What is cheating? What is plagiarism? What is the Honor Pledge?
- Code of Student Conduct: What behavior is prohibited?
- Sexual Misconduct: What to do in case of sexual harassment or sexual assault.
- Discrimination: Procedures to prohibit discrimination, complaints about discrimination, harassment, and retaliation.
- Accessibility: Information about disability support services (DSS) and accommodations.
- Student Rights Regarding Undergraduate Courses: What should I find in the course syllabus? Am I allowed to see my exams after they are graded?
- Official UMD Communication: Use of email, communication with faculty, communication about cancelled class meetings, and weather-related or other urgent notifications.
- Mid-Term Grades: Provided for 100 and 200 level courses, and all student athletes.
- Complaints about Course Final Grades: Questions about course grades should first be addressed to the course instructor.
- Copyright and Intellectual Property: Who owns the work that I produce in class?
- Course Evaluations: The School of Public Health is committed to the use of student course evaluations for improving the student experience, course and curriculum delivery, and faculty instruction.
- Campus Resources: ELMS, counseling, learning workshops, tutoring, writing help, questions about graduation, adding or dropping classes, withdrawing from the semester, etc.

Course Procedures and Policies:

Email: Verify your email address by going to www.my.umd.edu. Also make sure that this email account is linked to your Canvas account. All enrolled students are provided access to the University's email system and an email account. All official University email communication will be sent to this email address (or an alternate address if provided by the student).

Email has been adopted as the primary means for sending official communications to students, so email must be checked on a regular basis. Students are responsible for keeping their email address up to date or for redirecting or forwarding email to another address. Failure to check email, errors in forwarding email, and returned email (from "full mailbox" or "unknown user" errors for example), will not excuse a student from missing University announcement, messages, deadlines, etc.

Inclement Weather / University Closings / Emergency Procedures: In the event that the University has a delayed opening or is closed for an emergency or extended period of time, the instructor will communicate to students regarding schedule adjustments, including rescheduling of examinations and assignments due to inclement weather and campus emergencies.

Grading Procedures:

70% Homework Assignments (each equally weighted)

30% Final Exam

Final letter grades will be assigned according to the following:

A+ = 98+

A = 94-97.999

A- = 90-93.999

B+ = 88-89.999

B = 84-87.999

B- = 80-83.999

C+ = 78-79.999

C = 74-77.999

C- = 70-73.999

D+ = 68-69.999

D = 64-67.999

D- = 60-63.999

F = below 60

Course Outline / Calendar:

Module	Date	Topic	Assignments
#1	7/10/18	Overview of Course	HW #1 Assigned (due 7/17/18)
#2	7/12/18	Features of Complex Survey Data	
#3	7/17/18	Descriptive Statistics for Continuous Variables	HW #2 Assigned (due 7/24/18)
#4	7/19/18	Categorical Variable Analysis	
#5	7/24/18	Linear Regression with Survey Data – Part I	HW #3 Assigned (due 7/31/18)
#6	7/26/18	Linear Regression with Survey Data – Part II	
#7	7/31/18	Logistic Regression with Survey Data – Part I	HW #4 Assigned (due 8/7/18)
#8	8/2/18	Logistic Regression with Survey Data – Part II	
#9	8/7/18	Survival Analysis with Survey Data	HW #5 Assigned (due 8/14/18)
#10	8/9/18	Domain Estimation	
	8/14/18	*** Online Final Exam Made Available ***	Due 8/19/18

Detailed Outline of Modules	
Module #1: Overview of Course	7/10/18
<p>Topics: Outline of the Course, Logistics, Discussion of Homework and Exams, Brief Review and Recommended Resources for SAS</p> <p>HW #1 assigned – covers Modules #1 and #2 and is due no later than 11:59 p.m. on Tuesday 7/18/18</p>	
Module #2: Features of Complex Survey Data	7/12/18
<p>Topics: Description and Examples of Complex Survey Data Features (i.e., Finite Population Correction Factor, Stratification, Clustering, Weighting), Design Effects, Ultimate Cluster Assumption, Overview of the SURVEY family of SAS Procedures</p> <p>Required reading: Chapter 1</p> <p>Recommended reading: Chapter 2</p>	
Module #3: Descriptive Statistics for Continuous Variables	7/17/18
<p>Topics: Estimating Means, Totals, Ratios, and Quantiles</p> <p>Featured SAS procedure: PROC SURVEYMEANS</p> <p>Required reading: Chapter 3</p> <p>HW #2 assigned – covers Modules #3 and #4 and is due no later than 11:59 p.m. on Tuesday 7/24/18</p>	
Module #4: Categorical Variable Analysis	7/19/18
<p>Topics: Alternative Confidence Interval Methods for Extreme Proportions, Relative Risk Statistics, Odds Ratios, Chi-Square Tests</p> <p>Featured SAS procedure: PROC SURVEYFREQ</p> <p>Required reading: Chapter 4</p>	
Module #5: Linear Regression with Survey Data – Part I	7/24/18
<p>Topics: Review of Traditional Theory and Interpretation of Linear Regression Models, Matrix Notation and Estimation, Parameterizing Categorical Effects, Hypothesis Testing of Parameters</p> <p>Required reading: Chapter 5, Sections 5.1 – 5.2</p> <p>HW #3 assigned – covers Modules #5 and #6 and is due no later than 11:59 p.m. on Tuesday 7/31/18</p>	
Module #6: Linear Regression with Survey Data – Part II	7/26/18
<p>Topics: Extensions of Linear Regression to Handle Complex Survey Data, Weighted Least Squares, Adjusted F Statistic Testing for a Reduced Model</p> <p>Featured SAS procedure: PROC SURVEYREG</p>	

Required reading: Chapter 5, Sections 5.3 – 5.6	
Module #7: Logistic Regression with Survey Data – Part I	7/31/18
<p>Topics: Motivation and Advantages of the Logit Transformation, Review of Traditional Theory and Interpretation of Logistic Regression Models, Estimation and Hypothesis Testing of Parameters</p> <p>Required reading: Chapter 6, Sections 6.1 – 6.2</p> <p>HW #4 assigned – covers Modules #7 and #8 and is due no later than 11:59 p.m. on Tuesday 8/7/18</p>	
Module #8: Logistic Regression with Survey Data – Part II	8/2/19
<p>Topics: Extensions of Logistic Regression to Handle Complex Survey Data, Pseudo-Maximum Likelihood Parameter Estimation, Adjusted F Statistic Testing for a Reduced Model, Customizing Odds Ratios, Extensions to Cumulative Logit and Generalized Logit Models</p> <p>Featured SAS procedure: PROC SURVEYLOGISTIC</p> <p>Required reading: Chapter 6, Sections 6.3 – 6.8</p>	
Module #9: Survival Analysis with Survey Data	8/7/18
<p>Topics: Data Collection Strategies, Censoring Mechanisms, Classification of Survival Analysis Models, Plotting Survival/Hazard Functions, Cox Proportional Hazards Regression Models, Discrete-Time Hazards Regression Models</p> <p>Featured SAS procedures: PROC SURVEYPHREG and PROC SURVEYLOGISTIC</p> <p>Required reading: Chapter 7</p> <p>HW #5 assigned – covers Modules #9 and #10 and is due no later than 11:59 p.m. on Tuesday 8/14/18</p>	
Module #10: Domain Estimation	8/9/18
<p>Topics: Proper vs. Non-Proper Subsets, Domain-Specific Weights, Significance Testing of Domain Means, Degrees of Freedom Computations</p> <p>Required reading: Chapter 8</p>	

Additional Resources:

Another excellent textbook covering much of the same material as the required text for this course is *Applied Survey Data Analysis* by Heeringa, West, and Berglund, now in its second edition. Although the book primarily features examples using Stata, the authors developed syntax files free to download from their book website (<http://www.isr.umich.edu/src/smp/asda/>) that accomplish many of the same analytic tasks using SAS, R, SPSS, SUDAAN, WesVar, IVEware, and MPLUS.

And while a trove of documentation and user group conference proceedings is available online for free, students seeking a more formal resource for an introduction to SAS might consider obtaining one of these books:

- Delwiche, L., and Slaughter, S. (2013). *The Little SAS® Book: A Primer. Fifth Edition*. Cary, NC: SAS Institute.
- Li, A. (2013). *Handbook of SAS® DATA Step Programming*. Boca Raton, FL: CRC Press.